Informative Counting: Fine-grained Batch Authentication for Large-Scale RFID Systems

Wei Gong, Kebin Liu, Xin Miao, Qiang Ma, Zheng Yang, Yunhao Liu
School of Software, TNLIST, Tsinghua University
{gongwei, kebin, miao, maq, yang, yunhao}@greenorbs.com

ABSTRACT

Many algorithms have been introduced to deterministically authenticate Radio Frequency Identification (RFID) tags, while little work has been done to address the scalability issue in batch authentications. Deterministic approaches verify tags one by one, and the communication overhead and time cost grow linearly with increasing size of tags. We design a fine-grained batch authentication scheme, INformative Counting (INC), which achieves sublinear authentication time and communication cost in batch verifications. INC also provides authentication results with accurate estimates of the number of counterfeiting tags and genuine tags, while previous batch authentication methods merely provide 0/1 results indicating the existence of counterfeits. We conduct detailed theoretical analysis and extensive experiments to examine this design and the results show that INC significantly outperforms previous work in terms of effectiveness and efficiency.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; C.2.1 [**Network Architecture and Design**]: Wireless Communication

Keywords

RFID tags, batch authentication, informative counting

1. INTRODUCTION

Counterfeiting products are growing dramatically in terms of quantity, sophistication, category of goods, and countries affected in the late years [1]. The Counterfeiting Intelligence Bureau of the International Chamber of Commerce estimates that the share of counterfeiting commodities in international trade is about 5% to 7% [2]. The overall economic loss around the world amounts to more than $600 billion and is growing steadily over years [3]. Radio Frequency Identification (RFID) is one of the most promising technologies to help distinguish the genuines from the counterfeits. The RFID technology has several advantages over traditional methods , e.g. bar codes. First, the tag can be read inside containers and covers. Second, the identification information on the tag is unique for each affixed object. Of greater importance, hundreds of tags can be read at a time while bar codes can only be read one by one.

The problem of authenticating tags in large-scale RFID systems can be easily reduced to verifying each tag one by one. A number of different authentication protocols have been proposed to address this issue. Let N be the number of tagsIDs stored in the authentication server. Weis et al. [4] introduce a hash-based approach named Hash Lock. The search complexity of this method is $\mathcal{O}(N)$. In order to optimize searching efficiency, several tree-based approaches are proposed [5][6]. While tree-based data structures effectively reduce search complexity to $\mathcal{O}(\log N)$, they also additionally associate each tag with $\mathcal{O}(\log N)$ items in trees. By reviewing previous methods, we discover three major challenging issues affecting the effectiveness and efficiency of batch authentication in RFID system. First, the scanning time is not scalable. As all tags share communication channel, the reader has to introduce proper anti-collision scheme to receive authentication information from different tags. As a result, the scanning time is $\mathcal{O}(n)$, where n is the number of tags to be verified. Second, the communication overhead is not scalable. In those tree-based approaches, $\mathcal{O}(logN)$ hash values are expected to be exchanged between reader and server for each tag. Thus, the overall communication cost is $\mathcal{O}(n \log N)$. Third, the result of previous batch authentication is coarse-grain. The most recent work SEBA [7] can only provide 0/1 authentication result with probabilistic guarantees for a batch of tags. In particular, it only tells whether there exist counterfeits in verified tags, no further information.

In practical RFID systems, for those very expensive brands and valuable objects such as diamonds and art works, high cost of the deterministic authentication is worthwhile. However, it may not suit well for large amounts of fast-moving consumer goods such as fashion clothes, accessories and wines, which are the leading industries significantly affected by counterfeits. As a matter of fact, authenticating each tag of large quantities of fast-moving consumer goods is not necessary. Instead, knowing the approximate count of counterfeits and genuines with accuracy and error probability guarantees is desired in many large-scale RFID system applications.

MobiHoc'13, July 29–August 1, 2013, Bangalore, India.

In this paper, we propose a fine-grained batch authentication scheme, INformative Counting (INC), which provides a scalable and reliable batch authentication solution for large-scale RFID system. In particular, we geometrically divide the tag set and construct its Authentication Synopsis (AS). With the help of AS, we develop authentication algorithms to approximate both the number of counterfeits and genuines for any given (ε, δ) requirement. Based on authentication algorithms, fine-grained authentication protocols are provided to support fast batch authentication in practical large-scale RFID systems. To the best of our knowledge, we are the first to propose fine-grained batch authentication scheme for large-scale RFID systems. The major contributions of this work are as follows.

1. We propose a fine-grained batch authentication scheme INC. Compared with previous methods that produce only 0/1 authentication results, the result of INC additionally consists of accurate approximations of both the number of counterfeits and genuines in tested tags.

2. Using AS data structure, INC achieves sublinear authentication time and sublinear communication overhead in batch operation while existing work are linear with the number of tags.

3. We validate the proposed algorithms through theoretical analysis and conduct extensive simulations to verify the effectiveness and performance of our scheme.

The rest of this paper is organized as follows. Section 2 discusses related work. We introduce system model in Section 3. The INC design and theoretical analysis are presented in Section 4. We have more detailed discussions on this design in Section 5, and then show simulations results in Section 6. The conclusion is in Section 7.

2. RELATED WORK

Identifying counterfeiting goods serves the main purpose of the RFID tag usage. In early work, much of interests is centered on how to identify a single tag in deterministic and secure ways. A number of hardware-based approaches and security protocols are designed for system anonymity and anti-cloning [8][9][10]. As wireless medium is shared among all tags, the problem of resolving collisions arises when operating a batch of tags. Most of anti-collision schemes can be classified into two categories: ALOHA-based [11][12][13][14][15] and tree-based [16][17]. ALOHA-based protocols have been implemented in EPCGlobal Generation-2 RFID standard [18] and many commercial RFID system. Sheng et al. [12] introduce an efficient continuous scanning scheme which effectively uses the information collected in the previous scanning. While another popular standard, ISO 18000-6 [19], adopts tree-based schemes. The collision resolving processes are dramatically boosted, since all tags are itemized into query tree according to *tagID*s. Recently many methods aiming at providing private authentication are proposed. Weis et al. [4] propose Hash Lock scheme to protect tags from being attacked. But its searching complexity is linear with the size of keys in database. In order to speed up searching process, tree-based data structures are introduced to achieve logarithmic scale in [5][6][20]. Lu et al. even achieve $\mathcal{O}(1)$ authentication efficiency for a single tag, based on their own weak privacy model [21]. However, even combing state-of-the-art anti-collision methods and tree-based authentication schemes, the maximum identification throughput is still linear with the number of tags. Therefore they do not scale well when the size of tags quickly increases.

Besides above deterministic approaches, several probabilistic schemes are proposed to efficiently estimate the cardinality of tags. Kodialam et al. [22] propose Unified Simple Estimator (USE) and Unified Probabilistic Estimator (UPE) using linear counting technique. Qian et al. [23] introduce geometric distribution hashes to quickly estimate the cardinality of tags and the proposed LOF algorithm achieves $\mathcal{O}(\log n)$ time complexity. Zheng et al [24] propose Probabilistic Estimation Tree (PET), which advances estimation efficiency to $\mathcal{O}(\log \log n)$. A new scheme, Average Run based Tag estimation (ART), is 7x faster than UPE in the most recent work [25]. Li et al. [26] first propose energy-efficient RFID estimation algorithms for active tags. Zheng et al propose Zero-One Estimator (ZOE) protocol which rapidly converges to optimal parameter settings and achieves high estimation efficiency [27]. Although those schemes can estimate the number of distinct tags in RFID systems, they do not discriminate genuine ones from counterfeit ones.

In [7], a batch authentication scheme SEBA is proposed. It is able to detect counterfeits with probabilistic guarantee if the percentage of counterfeit tags is above predefined threshold. Nevertheless, SEBA has several drawbacks. First, its result is binary, merely indicating the existence of counterfeits in batch tags. Second, the frame size is $\mathcal{O}(N)$ and so it is not scalable in large-scale RFID systems. In contrast, INC can authenticate a batch of tags with accurate estimates of the number of counterfeits and genuines. And the authentication time and communication overhead of INC are sublinear with the cardinality of tags.

3. SYSTEM MODEL

In our system model, an RFID system consists of three main parts: one or more servers, several readers and hundreds of tags. Each tag is associated with a unique key, or called *tagID*. And it is attached to the object as an exclusive identity. Through wireless access medium, the reader can interrogate and receive responses from tags. The server usually plays a role of managing all keys of tags, including creation, authentication, and revocation of keys. If the key of a tag is actually stored in the server, we call this tag is *genuine*, otherwise *counterfeit*. Similar to most prior tag authentication schemes, *we do not discuss the issue that the genuine tag is attached to counterfeiting goods, vice versa*. The reader connects to the server through high speed wire or wireless networks. Let N be the number of keys maintained in the server, and n be the cardinality of batch tags to be verfied.

We adopt Listen-before-Talk [28] as the communication model between tag and reader in which the tag listens to the reader's interrogation and then replies. We also assume framed slotted ALOHA model as in [11][12][29][30]. In ALOHA model, the reader first broadcasts frame size f to all tags. Then each tag generates hash value $h(tagID)$ as its slot number. The reader then initializes time slot by sending "slot start" command. If the tag's slot number equals zero, it replies a bit-string to reader, otherwise it decreases slot number by one. This process repeats until f time slots are finished. In addition, we assume the bit-string contains some

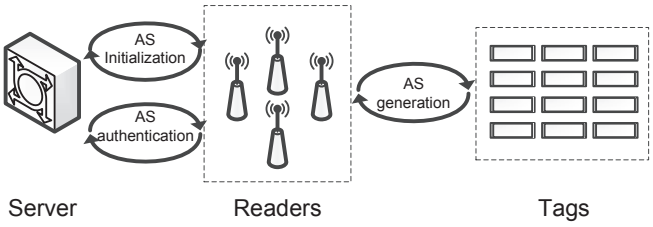


Figure 1: System architecture.

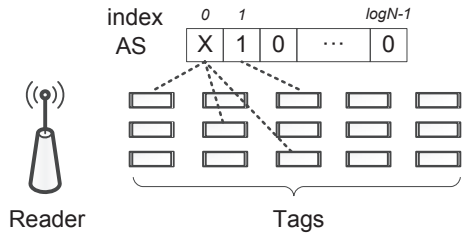


Figure 2: AS construction.

error-detecting codes like CRC. Note that this bit-string is not the identification information for each tag and is just used to detect whether there exist collisions in the time slot. Therefore, 10 bits should be fairly enough [22]. We classify time slots into three categories: *zero slot* means that there is no transmission in this slot while *singleton slot* denotes that only one tag transmits in this slot, and *collision slot* indicates that collisions happen in this slot.

There are some essential requirements of a good batch authentication scheme for large-scale RFID systems. First, the authentication scheme should be efficient, i.e., the authentication time and communication overhead should sublinearly or near-constantly increase with the cardinality of tags. Second, the authentication result should be informative to support various application demands. Knowing that whether there exist counterfeits in a batch of tags or not is far from adequate, since the administrator of RFID systems may still resort to per-tag authentication to count how many counterfeits in a number of tags. The per-tag verification is rather time-consuming for batch processing [7]. Third, if the authentication scheme can output estimates of the number of counterfeits and genuines, these estimates should be arbitrarily accurate. The last but not the least, the frame size in batch authentication cannot be infinitely large. As stated in [18], the frame size should be set no more than 512 for practical reasons.

We use two parameters as accuracy requirements of the estimation result: relative error ε between 0 and 1, and error probability δ between 0 and 1. Let n_c to be the actual number of counterfeits in a batch of tags, the output result $\hat{n_c}$ of an estimator with required (ε, δ) should satisfy $Pr[|n_c - \hat{n_c}| \leq \varepsilon n_c] \geq 1 - \delta$. For example, if the exact number of counterfeits in batch tags is 1000, and $\varepsilon = 0.05$, $\delta = 0.05$, then the output estimate is between 950 and 1050 with probability at least 0.95. In our design, these two parameters can be arbitrarily small.

4. THE DESIGN OF INFORMATIVE COUNTING

In this section, we first outline the framework of our authentication system in Section 4.1. Then we present how to generate authentication synopsis between reader and tag in Section 4.2. Authentication algorithms of estimating the number of counterfeits and genuines are detailed in Section 4.3.

4.1 System Architecture

The INC authentication scheme consists of three steps: AS initialization, AS generation and AS authentication. As shown in Figure 1, during AS initialization phase, after acquiring the accuracy requirement (ε, δ) from the user, the reader sends authentication request to the server, e.g., the number of hash functions. The server replies authentication parameters to the reader. Then in the AS generation phase, the reader interrogates batch tags and waits for responses. Through several rounds, collected ASes are transmitted to the server for the next AS authentication step. The server runs informative counting algorithms to give authentication results with accurate estimates of the number of counterfeits and genuines.

4.2 Authentication Synopsis

Our AS data structure is an extension of FM-Sketch [31]. In particular, as shown in Figure 2, the first position marked as "X" that denotes collisions happened in this time slot. The second position marked as "1" indicates there is only one tag transmitting in this slot. And the third position marked as "0" denotes that there is no response in this slot. The length of AS is equal to the frame size f. Similar to FM-Sketch, we do assume an ideal uniformly random hash function h_u generating hash values between 0 and $2^f - 1$. Then the second hash H, which counts the number of leading zeros (leftmost) of former hash value, is used to acquire geometric distribution as in [23]. It is expected that there are $\frac{n}{2^t}$ tags responses in the t-th position. For example, if the frame size is 8, one tag with $H(h_u(tagID_1)) = H(7) = H(00000111)_2 = 5$ should reply at slot 5. According to the geometric distribution characteristic, the frame size is $\log N$. So the frame of size 32 can support RFID system of which the cardinality is less than $4,294,967,295$. Thus, in this paper, we set the frame size at 32 which is fairly large for practical use.

The pseudocode of AS generation algorithms for tag and reader are given in Algorithm 1 and Algorithm 2 respectively. In Algorithm 1, the tag will generate the reply slot number k according to the two hashes H and h after receiving the probing message which might contain frame size and random seed. When each time slot starts, the tag responds instantly to the reader if its k is equal to 0. Otherwise it decreases k by 1 and keep silent. In Algorithm 2, the reader first broadcasts a request to all tags. And then the reader listens the status of tag responses in each time slot and sets the flag of slot according to different type of tag responses (0,1,X).

4.3 Algorithms

As in Figure 3, we assume that S is the set of tags on the server, and T is the set of tags to be tested. Accordingly, M is the union of S and T, i.e., $S \cup T$. C is the set of counterfeits which are in T but not in S, i.e., $T - S$. Therefore, we can easily define the number of counterfeits

Algorithm 1 AS generation algorithm for the tag

```
1: Receive a probing message from reader, compute slot
   number k = H(h_u(tagID)).
2: while TRUE do
3:    wait-for-slot-start().
4:    if k == 0 then
5:       respond instantly.
6:    else
7:       k ← k − 1, keep silent.
8:    end if
9: end while
```

Algorithm 2 AS generation algorithm for the reader

```
1: Initialize AS[i] ← 0(0 ≤ i ≤ log N − 1);
2: Broadcast a request to tags.
3: for i = 0 to log N − 1 do
4:    wait-for-tags-response().
5:    if there is no response in this slot then
6:       AS[i] ← 0.
7:    else
8:       if there is one tag response in this slot then
9:          AS[i] ← 1.
10:      else
11:         AS[i] ← X.
12:      end if
13:   end if
14: end for
```

as $|C| = |T - S|$. Similarly, we use G to denote the set of genuines which are both in S and T , i.e., $T \cap S$. And the number of genuines is denoted as $|G| = |T \cap S|$. If there is no counterfeits in the tested tags, then $M = \emptyset$ and $|C| = 0$. The basic idea of our approach is to get an (ε, δ) approximation to $|M|$, and then estimate the PG which is the $|G|$ to $|M|$ ratio and PC which is that $|C|$ to $|M|$. Finally, combining the accurate approximation of $|M|$ and above two ratios, the (ε, δ) approximations for both $|G|$ and $|C|$ are deduced. Note that as these two ratios are constants during the authentication process, custom probabilistic algorithms can be introduced to provide arbitrarily accurate estimates. To simplify the exposition, we treat δ as $\Theta(\delta)$ in following algorithm descriptions.

4.3.1 Estimating $|M| = |S \cup T|$

Estimating $|M|$ mainly consists of four steps. First, the server needs to collect enough number of independent AS_i^T from the reader and then generates its own $AS_i^S (1 \le i \le d$ using the same hash function respectively. Second, we construct virtual ASes slot by slot based on the criteria that only if both the corresponding slots of AS_i^T and AS_i^S are empty slots, the slot of virtual ASes would be empty. Otherwise, it is non-empty slot. Third, we find the smallest index level r which satisfying the appropriate threshold and obtain the non-empty probability of M of this level, p_r. Finally, as the expectation of this probability is the function of $|M|$, inverting of this function provides a good estimate, $\hat{M}$.

As shown in Figure 4(a), T_i is an AS from the reader and S_i is the corresponding AS generated by the server. 0 denotes empty slot and 1 denotes non-empty slot. The virtual AS construction is to combine non-empty slots. The slot of

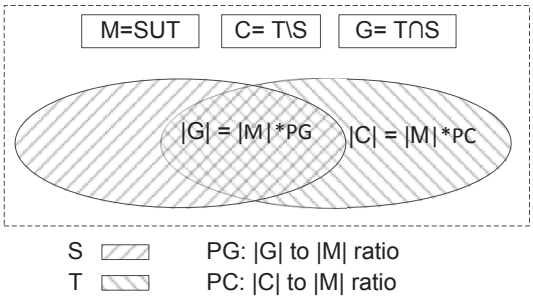


Figure 3: Basic idea of estimating the numbers of counterfeits and genuines.

virtual AS V_i would be non-empty if any one corresponding slot of S_i or T_i is non-empty. The index level search process is in Figure 4(b), NCount is the count of non-empty slots in specific index level and assumed threshold is 3. Therefore, starting from lowest index level (leftmost), the first index level of which the NCount is below the threshold would be the appropriate estimating level. And here we find the qualified level is the third index.

Definition 1. For a specific index r $(0 \le r \le 31)$. We define

$$x(r) = \begin{cases} 1 & \text{the index } r \text{ of } AS^{S \cup T} \text{ is non-empty} \\ 0 & \text{otherwise} \end{cases}$$

Therefore the $x(r)$ takes 1 with probability $p_r = 1 - (1 - \frac{1}{U_r})^{|M|}$, where $U_r = 2^{r+1}$.

From above definition, we can invert the probability formula of p_r to deduce the estimate $\hat{M}$

$$\hat{M} = \frac{\ln(1 - \hat{p_r})}{\ln(1 - \frac{1}{U_r})}, \tag{1}$$

where $\hat{p_r}$ is the observed non-empty probability of index level r.

Now, we have the $|M|$ estimating function but there are still two major problems. 1. d_m. How many pairs of independent ASes from reader and server are enough to provide an accurate estimate $\hat{p_r}$? 2. λ. How accurate of $\hat{p_r}$ is adequate to produce the user-specified (ε, δ) approximation of $|M|$? The following lemmas answer those questions.

LEMMA 1. *If $d_m \ge \frac{96}{7}\lambda^{-2} \ln \frac{2}{\delta}$, r is the smallest index satisfying $NCount_r \le \frac{(1+\lambda)d_m}{4}$, and $\hat{p_r} = \frac{NCount_r}{d_m}$, then $Pr[|\hat{p_r} - p_r| \le \lambda p_r] \ge 1 - \delta$.*

PROOF. Note that the probability that the index r is non-empty in $AS^{S \cup T}$ is the same as the probability that the index r is non-empty in either AS^S or AS^T. So we fix r to a positive value such that $\frac{1}{8} \le \frac{M}{U_r} \le \frac{1}{4}$. By binomial expansion, we know that $p_r = 1 - (1 - \frac{1}{U_r})^M = \sum_{i=1}^{M} (-1)^{i+1} \binom{M}{i} U_r^{-i}$, then $(\frac{M}{U_r} - \frac{1}{2}(\frac{M}{U_r})^2) \le p_r \le \frac{M}{U_r}$. Therefore, we obtain that $\frac{7}{32} \le p_r \le \frac{1}{4}$. Also remember that $x(r)$ is a binomial random variable, hence we can apply Chernoff bound [32]. Using slightly worse case bound expressions, we know that as long as $d_m p_r \ge \frac{3}{\lambda^2} \ln \frac{2}{\delta}$, i.e,

$$d_m \ge \frac{96}{7\lambda^2} \ln \frac{2}{\delta}, \tag{2}$$

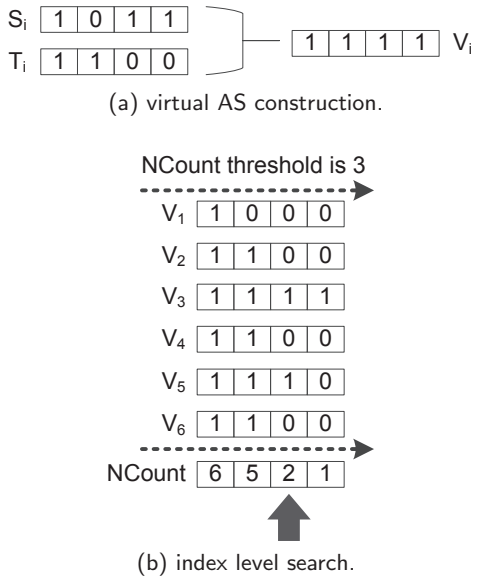


Figure 4: An illustration of estimating $|M| = |S \cup T|$.

as $p_r \geq \frac{7}{32}$, then the estimate $\hat{p_r} = \frac{NCount_r}{d_m}$ can satisfy $Pr[|\hat{p_r} - p_r| \leq \lambda p_r] \geq 1 - \delta$. Therefore at index level r, we can find $|\hat{p_r} - p_r| \leq \lambda p_r$ which ensures that $count_r \leq \frac{(1+\lambda)d_m}{4}$ with probability at least $1 - \delta$. □

LEMMA 2. *If $p_r < \frac{1}{4}$, $\lambda = \frac{\varepsilon}{2}$ and $|\hat{p_r} - p_r| \leq \lambda p_r$, then a new estimate $\hat{M}$, defined as $\hat{M} = \frac{\ln(1-\hat{p_r})}{\ln(1-\frac{1}{U_r})}$, satisfies $|\hat{M} - M| \leq \varepsilon M$.*

PROOF. For a continuous function $f(x) = \ln(1-x)$, we know that $\ln(1-x) < -x$ when $0 < x < 1$. And also there is $|f(x) - f(\bar{x})| \leq \varepsilon |sup_{y \in (x,\bar{x})} f\prime(y)|$ if $\bar{x}$ is close to x, hence we can get $|\ln(1-x) - \ln(1-\bar{x})| \leq \frac{|x-\bar{x}|}{1-max\{x,\bar{x}\}}$. Therefore,

$$\begin{aligned}
|\hat{M} - M| &= \frac{|\ln(1-p_r) - \ln(1-\hat{p_r})|}{-\ln(1-\frac{1}{U_r})} \\
&\leq -\ln(1-\frac{1}{U_r}) \cdot \frac{|p_r - \hat{p_r}|}{1 - max\{p_r, \hat{p_r}\}} \\
&\leq -\ln(1-\frac{1}{U_r}) \cdot \frac{\frac{\varepsilon}{2} p_r}{1 - (1+\frac{\varepsilon}{2}) p_r} \\
&\leq -\ln(1-\frac{1}{U_r}) \cdot \varepsilon p_r \\
&\leq -\ln(1-\frac{1}{U_r}) \cdot \varepsilon \cdot (-\ln(1-p_r)) \\
&= \varepsilon M.
\end{aligned}$$

□

Combining lemma 1 and lemma 2, we can establish following theorem.

THEOREM 1 ($|M|$ ESTIMATE). *If $d_m \geq \frac{384}{7\varepsilon^2} \ln \frac{2}{\delta}$, equation 1 outputs an (ε, δ) estimate $\hat{M}$ for $|M|$.*

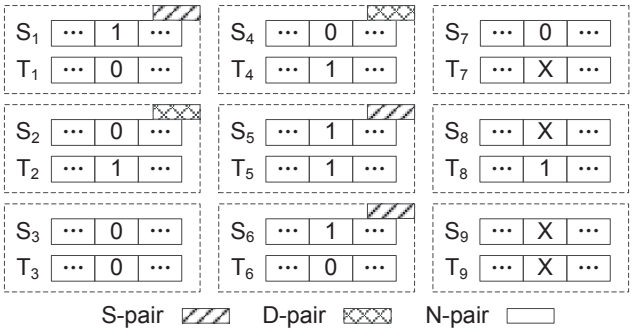


Figure 5: Distinguish different types of pairs for estimating the number of counterfeits.

4.3.2 *Estimating the number of counterfeits*

Given the set S and T, the cardinality of the set-difference $|C| = |T - S|$ is the number of distinct tags that are in T but not in S, i.e., the counterfeit ones. Here, we introduce an (ε, δ) approximation scheme for estimating $|C|$ based on the former accurate estimate $\hat{M}$. It is worth noting that a "naive" method may be proposed to use $|T - S| = |T \cup S| - |S|$ formula and corresponding (ε, δ) estimates of both $|T \cup S|$ and $|S|$ to estimate the cardinality of this set-difference. However, this "naive" approach does not give any accuracy guarantees of the approximation for $|T - S|$, i.e., it is not an (ε, δ) estimation scheme as needed.

Estimating $|C|$ procedure is composed of three essential parts. First, assume that we get an estimate $\hat{M}$ with $(\frac{\varepsilon}{3}, \delta)$. And then we choose the level $r_c = \lceil \log(\frac{\alpha \hat{M}}{1-\varepsilon}) \rceil$, such that the number of elements that can be hashed into this level, 2^{r_c+1}, is slightly greater than $\hat{M}$, where α is a constant parameter greater than 1 (line 3). Second, in order to obtain difference ratio (p_d), we seek to distinguish different types of pairs and record their counts. Third, we combine the observed difference ratio and the estimate $\hat{M}$ to compute the approximation of set difference $\hat{C}$.

The core of estimating $|C|$ is the second part. As shown in Figure 5, there are 9 AS pairs. And if one slot is "1" and the corresponding slot is "0" or "1", the pair is called singleton pair, S-pair for short. Further if the slot of S_i is '0' and slot T_i is '1', the pair is denoted as difference pair, D-pair for short. The other pairs are called N-pair. Obviously, all D-pairs are S-pairs. And the count of S-pair is 5 (singletonCount) and its of D-pair is 3 (diffCount). Therefore, we can roughly estimate that the number of counterfeiting ones is about $\frac{3}{5}|M|$.

Definition 2. For a specific index r_c ($0 \leq r_c \leq 31$), let $U_c = 2^{r_c+1}$, we define

$$x_M(r_c) = \begin{cases} 1 & \text{the index } r_c \text{ of } AS^{S \cup T} \text{ is singleton} \\ 0 & \text{otherwise.} \end{cases}$$

We can have that the $x_M(r_c)$ takes 1 with probability $p_M = |S \cup T| \cdot \frac{1}{U_c}(1 - \frac{1}{U_c})^{|S \cup T|-1}$, as the probability of a given element mapped to level r_c is $\frac{1}{U_c}$, and the probability of a given element being singleton is $\frac{1}{U_c}(1 - \frac{1}{U_c})^{|S \cup T|-1}$. And we also define

$$x_C(r_c) = \begin{cases} 1 & \text{the index } r_c \text{ of } AS^T \text{ is "1" and } AS^S \text{ is "0"} \\ 0 & \text{otherwise.} \end{cases}$$

Likewise, $x_C(r_c)$ takes 1 with probability $p_C = |T-S| \cdot \frac{1}{U_c}(1-\frac{1}{U_c})^{|S\cup T|-1}$. Therefore, we can define the conditional probability p_d as

$$p_d = \frac{p_C}{p_M} = \frac{|T-S| \cdot \frac{1}{U_c}(1-\frac{1}{U_c})^{|S\cup T|-1}}{|S\cup T| \cdot \frac{1}{U_c}(1-\frac{1}{U_c})^{|S\cup T|-1}} = \frac{|T-S|}{|S\cup T|}. \quad (3)$$

We also define $\acute{d_c} = \sum_{i=0}^{d_c-1} x_M(r)_i$ over d_c AS pairs, the number of S-pairs.

From above definitions, we can deduce the formula for estimating $|C|$ as

$$\hat{C} = \hat{p_d} \cdot \hat{M}, \quad (4)$$

where $\hat{p_d} = diffCount/singletonCount$.

In order to produce an (ε, δ) estimate $\hat{C}$, the following lemmas deal with three questions. 1. d_c. How many pairs of independent ASes are enough to produce specified number of S-pairs? 2. $\acute{d_c}$. How many S-pairs are needed to provide accurate estimate of p_d? 3. How accurate of $\hat{M}$ and $\hat{p_d}$ are adequate to give an (ε, δ) estimate $\hat{C}$?

LEMMA 3. *Let $\alpha > 1$ and $r_c = \lceil \log(\frac{\alpha|S\cup T|}{1-\varepsilon}) \rceil$. For any constant γ between 0 and 1, if $d_c \geq \frac{3\alpha^2}{\gamma^2(\alpha-1)} \ln\frac{2}{\delta}$, then $\acute{d_c} > \frac{(1-\gamma)(\alpha-1)}{\alpha^2} d_c$ with probability $1-\delta$.*

PROOF. As $r_c = \lceil \log(\frac{\alpha|S\cup T|}{1-\varepsilon}) \rceil$, we have that $\frac{|S\cup T|}{U_c} < \frac{1}{\alpha}$. By Bernoulli inequality [33], we can obtain the lower-bound of p_M

$$p_M > \frac{|S\cup T|}{U_c}(1 - \frac{|S\cup T|}{U_c}) > \frac{\alpha-1}{\alpha^2}.$$

Again by Chernoff bound [32], we can obtain an estimate $p_{\hat{M}} = \frac{\acute{d_c}}{d_c}$ such that $Pr[|p_{\hat{M}} - p_M| \leq \gamma p_M] \geq 1-\delta$ as long as $d_c \geq \frac{3\alpha^2}{\gamma^2(\alpha-1)} \ln\frac{2}{\delta} \geq \frac{1}{p_M}\frac{3}{\gamma^2}\ln\frac{2}{\delta}$. Thus

$$p_{\hat{M}} \geq (1-\gamma)p_M \Leftrightarrow \acute{d_c} \geq (1-\gamma)p_M \cdot d_c > \frac{(1-\gamma)(\alpha-1)}{\alpha^2} d_c.$$

□

LEMMA 4. *If $\acute{d_c} \geq \frac{|S\cup T|}{|T-S|}\frac{3}{\eta^2}\ln\frac{2}{\delta}$, $\hat{p_d}$ is an (η, δ) estimate.*

PROOF. By Chernoff bound [32], we know that as long as $\acute{d_c} p_d \geq \frac{3}{\eta^2}\ln\frac{2}{\delta}$, $\hat{p_d}$ is within a relative error η with probability $1-\delta$. And combing equation (3), it gives that $\acute{d_c} \geq \frac{|S\cup T|}{|T-S|}\frac{3}{\eta^2}\ln\frac{2}{\delta}$. □

LEMMA 5. *If we have an $(\frac{\varepsilon}{3}, \delta)$ estimate $\hat{M}$ for $|M|$, and an $(\frac{\varepsilon}{3}, \delta)$ estimate $\hat{p_d}$ for p_d, then $\hat{p_d}\hat{M}$ is an (ε, δ) estimate of $p_d M$, i.e., C.*

PROOF. $|\hat{p_d}\hat{M} - p_d M| = |p_d(1 \pm \frac{\varepsilon}{3})M(1 \pm \frac{\varepsilon}{3}) - p_d M| \leq p_d M(\pm\frac{2\varepsilon}{3} + \frac{\varepsilon^2}{9}) \leq \varepsilon p_d M$. □

By lemma 3 and 4, we know that in order to get an (η, δ) estimate $\hat{p_d}$, d_c should be at least $\frac{\alpha^2}{(1-\gamma)(\alpha-1)}\frac{|S\cup T|}{|T-S|}\frac{3}{\eta^2}\ln\frac{2}{\delta}$. And also by lemma 3, d_c should be at least $\frac{3\alpha^2}{\gamma^2(\alpha-1)}\ln\frac{2}{\delta}$. Combing those two conditions, hence we get

$$d_c \geq \frac{\alpha^2}{\min\{1-\gamma, \gamma^2\}(\alpha-1)}\frac{|S\cup T|}{|T-S|}\frac{3}{\eta^2}\ln\frac{2}{\delta} = d_c^1. \quad (5)$$

Algorithm 3 Estimating the number of counterfeits

Input: relative error ε, error probability δ.
Output: estimate $\hat{C}$.

compute d_c by theorem 2.
generate d_c independent AS pairs of S and T as in Algorithm 1 and 2.
$\hat{M}$=**GetEstimateM**($AS_{dc}, \frac{\varepsilon}{3}, \delta$).
$singletonCount \leftarrow 0$, $diffCount \leftarrow 0$.
$\alpha = 2$, $r_c \leftarrow \lceil \log(\frac{\alpha\hat{M}}{1-\varepsilon}) \rceil$.
for $i = 0$ to $d_c - 1$ **do**
  if ($AS_i^S[r_c]$ == 0 and $AS_i^T[r_c]$ == 1) or ($AS_i^S[r_c]$ == 1 and $AS_i^T[r_c]$ == 0) or ($AS_i^S[r_c]$ == 1 and $AS_i^T[r_c]$ == 1) **then**
    $singletonCount \leftarrow singletonCount + 1$.
    if $AS_i^S[r_c]$ == 0 and $AS_i^T[r_c]$ == 1 **then**
      $diffCount \leftarrow diffCount + 1$.
    end if
  end if
end for
return $\hat{C} = \frac{diffCount}{singletonCount} \cdot \hat{M}$.
GetEstimateM(AS_{dc},ε, δ)
$index \leftarrow 0$, $found \leftarrow false$, $\lambda \leftarrow \frac{\varepsilon}{2}$, $minC \leftarrow \frac{(1+\lambda)d_m}{4}$.
$d_m \leftarrow \frac{384}{7\varepsilon^2}\ln\frac{2}{\delta}$, randomly select d_m pairs AS_{dm} from AS_{dc}.
while $index \leq 31$ and $!found$ **do**
  $NCount \leftarrow 0$.
  for $i = 0$ to $d_m - 1$ **do**
    if $AS_i^S[index] \neq 0$ or $AS_i^T[index] \neq 0$ **then**
      $NCount \leftarrow NCount + 1$.
    end if
  end for
  if $NCount > minC$ **then**
    $index \leftarrow index + 1$.
  else
    $found \leftarrow true$.
  end if
end while
$\hat{p} \leftarrow \frac{NCount}{d_m}$, $\hat{M} \leftarrow \frac{\ln(1-\hat{p})}{\ln(1-\frac{1}{2^{index+1}})}$.
return $\hat{M}$.

In order to get minimum d_c, we can compute the optimal values for γ and α. The results are $\gamma = \frac{\sqrt{5}-1}{2}$, $\alpha = 2$. The first lower-bound of d_c is denoted as d_c^1. And by lemma 5, let $\eta = \frac{\varepsilon}{3}$, we can get an $(\frac{\varepsilon}{3}, \delta)$ estimate of p_d. By theorem 1, in order to get an $(\frac{\varepsilon}{3}, \delta)$ estimate for $|M|$, we get

$$d_c \geq \frac{3456}{7\varepsilon^2}\ln\frac{2}{\delta} = d_c^2. \quad (6)$$

The second lower-bound of d_c is denoted as d_c^2. Based on the above analysis, we can state the following theorem.

THEOREM 2 (ESTIMATING THE NUMBER OF COUNTERFEITS). *If $d_c = \max\{d_c^1, d_c^2\}$, equation 4 outputs an (ε, δ) estimate $\hat{C}$ for $|C|$.*

Algorithm 3 shows the pseudocode of estimating the number of counterfeits.

4.3.3 Estimating the number of genuines

According to the concept of genuineness, the number of genuines in T can be defined as $|G| = |S \cap T|$, i.e., the number of distinct keys that are in both T and S. Fortunately,

the (ε, δ) estimation algorithm of G is very similar to Algorithm 3. The only difference is that the condition of line 9 should be changed into "$AS_i^S[r_c] == 1$ and $AS_i^T[r_c] == 1$". Similarly, we can obtain the formula for estimating $|G|$ as

$$\hat{G} = \hat{p_s} \cdot \hat{M}, \tag{7}$$

where $\hat{p_s}$ is the identical ratio. Likewise, we derive the lower-bound of d_g as follows

$$d_g \geq \frac{\alpha^2}{\min\{1-\gamma, \gamma^2\}(\alpha-1)} \frac{|S \cup T|}{|S \cap T|} \frac{3}{\eta^2} \ln \frac{2}{\delta} = d_g^1 \tag{8}$$

$$d_g \geq \frac{3456}{7\varepsilon^2} \ln \frac{2}{\delta} = d_g^2 \tag{9}$$

Thus, we can establish the following theorem.

THEOREM 3 (ESTIMATING THE NUMBER OF GENUINES). *If $d_g = \max\{d_g^1, d_g^2\}$, equation 7 outputs an (ε, δ) estimate $\hat{G}$ for $|G|$.*

5. DISCUSSION

In this section, we discuss several important issues of our proposed algorithms.

5.1 t-wise Independent Hash Functions

So far, the design and analysis of our schemes assume that there exist ideal uniformly random hash functions. In fact this assumption is unrealistic for practical use. Therefore, we can employ t-wise independent hash functions to ensure our former analysis still hold. Actually, if $t = \Theta(\log(\varepsilon^{-1}))$, then using h_i from t-wise independent hash family $\mathcal{H}_t$ as alternatives of h_u in AS constructions, we can still provide (ε, δ) estimates for the number of counterfeits and genuines. Due to limited space, we omit the details here.

5.2 Singleton Slot Observation

In the line 5 of Algorithm 4, "$AS_i^S == 1$ and $AS_i^T == 1$" is one of conditions that singleton is found in $(S \cup T)$. This condition may fail if two elements are "luckily" enough to be mapped into the same slot. Fortunately, we can prove that the possibility of this "luck" is rather small and negligible in practical use. This possibility, denoted as p_l, is that two different elements are hashed into *index* level and also are singletons in their respective AS. Therefore, we have $p_l = \frac{1}{U_c}(1-\frac{1}{U_c})^{|S\cup T|-2} \approx \frac{1}{M}(1-\frac{1}{M})^{M-2}$ where $U_c = \frac{\alpha \hat{M}}{1-\varepsilon}$. Therefore, we know that even if the server only has $M = 10,000$ keys, then p_l is about $0.000,037$. And with the increasing number of keys on the server, p_l would be even smaller. As a matter of fact, we can also use an extra round to determine whether the undecided slot is the singleton slot in M using rolling schemes similar in [30].

5.3 The Size of Counterfeits/genuines

Another practical problem is that the size of counterfeits/genuines may be too small. For example, if $\frac{|S\cup T|}{|T-S|}$ is quite large, estimation problem would become difficult as in equation 5. To address this issue, we know that the share of counterfeiting commodities in real life is about 5% to 7% [2]. Therefore, we can provide the "sanity" lower bound $B = f(M, d_c, \varepsilon, \delta)$ determined by our theorem. Then our earlier statements about (ε, δ) estimate can be formed like this: "...outputs (ε, δ) estimate of $|C|$ as long as $|C| \geq B$".

Table 1: Scheme comparison

Scheme	Scanning Cost	Communication Cost	Authentication Cost
Hash Lock	$\mathcal{O}(n)$	$\mathcal{O}(n)$	$\mathcal{O}(nN)$
ACTION	$\mathcal{O}(n \log N)$	$\mathcal{O}(n \log N)$	$\mathcal{O}(n \log N)$
SEBA [1]	$\mathcal{O}(N)$	$\mathcal{O}(N)$	$\mathcal{O}(N)$
INC	$\mathcal{O}(\frac{N}{n} \log N)$	$\mathcal{O}(\frac{N}{n} \log N)$	$\mathcal{O}(\frac{N}{n} \log N)$

However, as shown later in evaluation part we know that this sanity bound is rather pessimistic, it may due to that fact that our analysis are based on the worst-case analysis. Furthermore, we may divide keys on the server into tree or other organized data structures in which the initial matching space $\mathcal{O}(N)$ might be effectively reduced. For example, keys of tagged items are group by different categories such as wine and clothes. And so when authenticating wines, S should be S_{wine}, not $S_{wine} \cup S_{clothes}$.

5.4 Time-efficient Optimization

Although the communication cost and AS construction time are sublinear with the cardinality of RFID system, there are possible solutions to further boost authentication efficiency. Remember in Algorithm 3 that the estimation are performed on some specific level, e.g., *index*, because the elements of ASes below this level are almost 'X' or '1'. Therefore, if we can obtain the scale of $|S \cup T|$ as a prior knowledge, AS can be compressed from $\log N$ slots into $(\log N - index)$ slots or $\log \log N$, where *index* is the estimation start level. Thus, compressed AS may significantly speed up the authentication process.

5.5 Energy-efficient Optimization

The energy cost of the tag is one important issue we should carefully cope with. For example, in a large warehouse equipped with RFID system, active tags are usually used to label commodities [26]. Since active tags are battery-powered, recharging batteries for thousands of tags is really a heavy work, and even in some cases the tags are not easily reachable. As our early analysis shown, although this number of ASes is sublinear with the cardinality of RFID system, it still imposes heavy burdens for resource constraint tags since all tags are required to respond to each interrogation from reader during authentication. To further reduce energy consumption of tags, mechanisms that may shift the energy consumption from tag side to reader side are necessary. One possible solution is to build an energy-efficient AS data structure of which the organization is based on a hash function (e.g., h_g) and each tag contributes to all ASes sequentially. The design this hash function is to divide tags into several groups and minimize the number of data transmissions of tags.

5.6 Comparison

Table 1 compares INC with three state-of-the-art authentication schemes: Hash lock [4], ACTION [6] and SEBA

[1] As optimal frame length functions are complex and implicit, therefore we use simple function as $\delta = 0.99$. See details in [7].

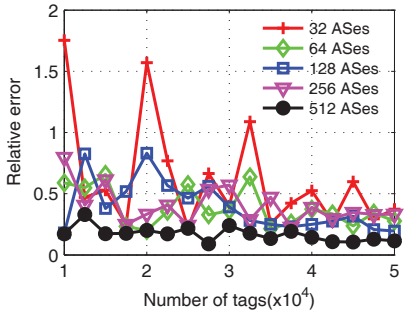


Figure 6: Relative error of estimate $\hat{C}$.

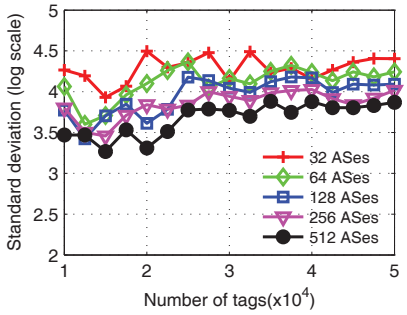


Figure 7: Standard deviation of estimate $\hat{C}$.

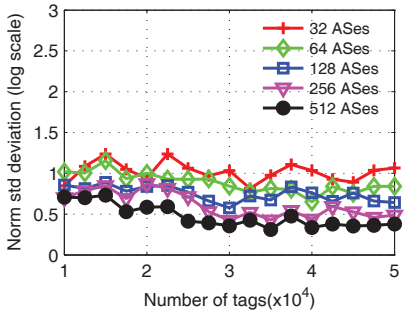


Figure 8: Normalized std deviation of estimate $\hat{C}$.

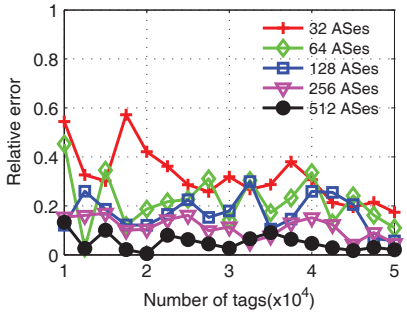


Figure 9: Relative error of estimate $\hat{G}$.

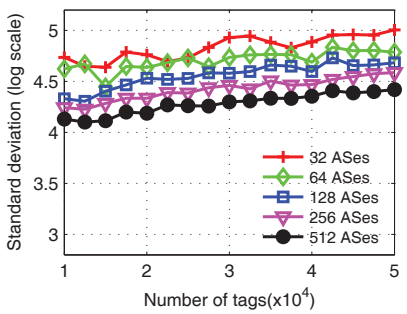


Figure 10: Standard deviation of estimate $\hat{G}$.

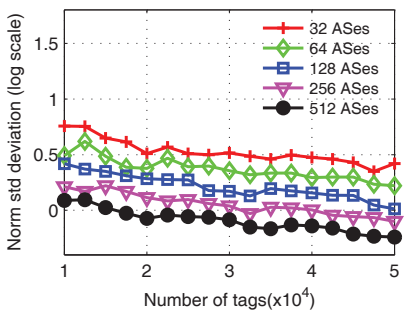


Figure 11: Normalized std deviation of estimate $\hat{G}$.

[7]. Refresh [21], which achieves better authentication efficiency over ACTION, is not included in this comparison in terms of fairness. Since it is based on the weak privacy model in which the improvement of the authentication efficiency is at the cost of the privacy degradation. The comparison mainly consists of three metrics: (i) scanning cost that is the complexity of acquiring the authentication information; (ii) communication cost which is the amount of data transmissions between reader and server; and (iii) authentication cost which is the time of verifying n tags in N server-stored keys. As Hash lock and ACTION are per-tag based deterministic authentication schemes, we assume an omniscient anti-collision solution is with them in which n tags can transmit identification information in a frame of size n perfectly. As characteristics of deterministic approaches, scanning cost and communication cost of Hash lock are $\mathcal{O}(n)$. With the help of tree-based data structures, ACTION achieves $\mathcal{O}(n\log N)$ efficiency in terms of scanning cost, communication cost and authentication cost, while authentication cost of Hash lock is $\mathcal{O}(nN)$ with linear searching. For SEBA [7], as the optimal frame length is $7(N + n\varepsilon)$, the scanning cost, communication cost and authentication cost are $\mathcal{O}(N)$. From theorem 2 and theorem 3, we know that for a given (ε, δ), the space complexity of ASes is $\mathcal{O}(\frac{\log(\delta^{-1})N}{\varepsilon^2 n}\log N)$. If we treat (ε, δ) as constants, therefore we can easily deduce the corresponding scanning cost, communication cost and authentication cost. We can see that if the size of tested set n is relatively small, ACTION maybe the best solution. But when n is growing larger and larger, INC wins the contest as its complexity is asymptotically towards $\mathcal{O}(\log N)$.

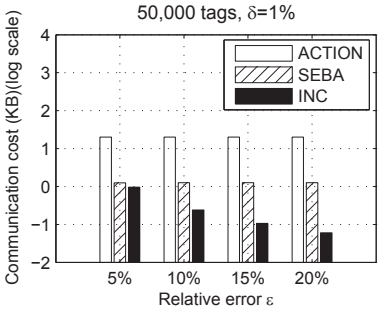
6. EVALUATION

We evaluate the performance of INC under extensive simulations. First, we study the estimation accuracy with tunable size of ASes under various settings. Then we compare INC with two most recent methods ACTION [6] and SEBA [7] in terms of scanning cost and communication overhead.

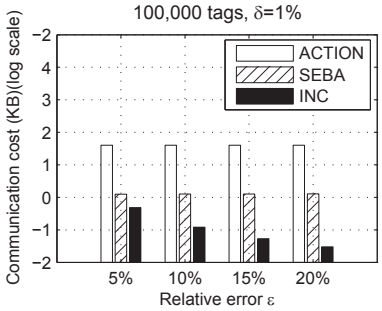
6.1 Setup and Metrics

The simulations are implemented on a laptop with Intel i7 CPU at 2.8GHz and 4GB RAM, using C# as programming language. In order to support ACTION of depth 21 with $N = 2^{20} = 1,408,576$ server tags, we have $2^{21} = 2,097,152$ keys stored in SQLExpress 9.00 database. For ACTION, we use MD5 as the uniform hash function. Therefore each tag is associated with 21 key values. According to reports in [2], we set the ratio of counterfeiting tags at 7%. We take 300 runs and report the average.

The estimation accuracy of the number of counterfeits or genuines is the most important metric for our authentication scheme. We use three standard parameters to measure the accuracy of INC. The first parameter is relative error: $RelError = |\frac{\hat{\theta}-\theta}{\theta}|$, where θ is the actual number and $\hat{\theta}$ is the estimate. The second parameter is standard deviation: $\sigma = \sqrt{E[(\hat{\theta}-\theta)^2]}$. The third parameter is normalized standard deviation: $\sigma_n = \frac{\sigma}{E[\theta]}$.



(a) with different relative error ε, 50,000 tags and $\delta = 1\%$.



(b) with different relative error ε, 100,000 tags and $\delta = 1\%$.

Figure 12: Communication cost between reader and server (in *log scale*).

Table 2: Total slots to meet $\delta = 1\%$ with different ε

ε	$n = 50,000$		
	ACTION	SEBA	INC
5%	1,050,000	1,051,076	808,992
10%	1,050,000	1,053,576	202,240
15%	1,050,000	1,056,076	89,888
20%	1,050,000	1,058,576	50,560
	$n = 100,000$		
5%	2,100,000	1,053,576	404,480
10%	2,100,000	1,058,576	101,120
15%	2,100,000	1,063,576	44,928
20%	2,100,000	1,068,576	25,280

6.2 INC Investigation

The results in Figures 6, 7, and 8, show that the accuracy of estimating the number of counterfeits in tested tags. From those figures, we make following observations. First, we can improve estimation accuracy of INC by increasing the number of ASes. As illustrated in Figure 6, increasing the number of ASs lead to significant drops of relative error. In particular, when the number of tags is 50,000, the relative error is 0.37 with 32 ASes and drops close to 0.1 with 512 ASes. This nice linearly addible property of INC provides us flexibility of making tradeoffs between estimation accuracy and efficiency according to different application demands. Second, the standard deviation can be greatly reduced by larger size of ASes, which is indicating in Figure 7. Third, normalized standard deviations are getting steadily smaller with increasing the cardinality of tags. As shown in Figure 8, the normalized standard deviation is 5.1 with 10,000 tags and diminishes to 2.4 with 50,000 tags using 512 ASes.

Similar trends can also be observed in Figures 9, 10 and 11. We can see that the estimate $\hat{G}$ of approximating the number of genuines is advantageous over $\hat{C}$ in all three aspects. As we stated in Section 5.3, the main reason is that the number of genuine tags is larger than its of counterfeits, since we set the share of counterfeiting tags at 7%. Relative errors with different size of tags from 10,000 to 50,000 with 512 ASes are below 0.2 and most of them are about 0.01. From normalized standard deviation perspective, $\hat{G}$ is also an excellent estimator. As shown in Figure 11, for 50,000 tags, the normalized standard deviation is 0.57 with only 512 ASes.

6.3 Performance Comparison

We compare the performance of INC with the two state-of-the-art approaches ACTION and SEBA. We mainly compare the scanning cost and communication cost under different accuracy settings. Here we measure scanning time in terms of the total time slots to acquire authentication information and communication cost in terms of data size of transmissions between reader and server.

We compare three methods given $\delta = 1\%$ and ε changing from 5% to 20% and the size of tags at 50,000 and 100,000. As shown in Table 2, INC significantly outperforms both ACTION and SEBA. For instance, total time slots of INC is 8.5% of SEBA when $\varepsilon = 15\%$, $n = 50,000$ and is 1.2% of ACTION when $\varepsilon = 20\%$, $n = 100,000$. The communication cost results of three schemes are depicted in Figure 12. We assume that the length of identification ID is 96 bits [18] in ACTION and the length of bit-string for each time slot is 10 bits [22] in SEBA and INC. Again, we can see that INC achieves much lower communication cost than both ACTION and SEBA, e.g., the data size of transmissions of INC is merely 0.3% of ACTION and 4.7% of SEBA when $\varepsilon = 20\%$, $n = 50,000$. From a different perspective, above comparison results indicate that given a certain amount of transmission data or scanning time requirement, the authentication accuracy of INC will be much better than ACTION and SEBA. As a matter of fact, since INC achieve $\mathcal{O}(\frac{N}{n} \log N)$ efficiency, when the size of tags quickly scales, the performance gain of INC over ACTION and SEBA is growing larger.

7. CONCLUSION

This paper proposes a probabilistic batch authentication schemes, INC, for large-scale RFID systems. Compared with previous methods. INC not only achieves sublinear authentication efficiency, but also provides accurate estimate of the number of counterfeits and genuines. Both theoretical analysis and extensive simulations are presented to show advantages of INC over prior work. In future work, we plan to examine whether our estimation bounds are tight. And

also we intend to extend our framework to multiple readers scenarios, in which each reader has its own operation range and tag set.

8. ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers for valuable and insightful comments. This work is supported in part by the NSFC Young Scholar 61103187, National Basic Research Program of China (973) Grant No. 2011CB302705, NSFC Major Program 61190110 and the NSFC under Grant 61171067. We also acknowledge the support from the codes of USRP2reader from the Open RFID Lab (ORL) project [34].

9. REFERENCES

[1] Thorsten Staake, Frĺędĺęric Thiesse, and Elgar Fleisch. Business strategies in the counterfeit market. *Journal of Business Research*, 65(5):658 – 665, 2012.

[2] ICC Counterfeiting Intelligence Bureau. Countering counterfeiting: A guide to protecting and enforcing intellectual property rights. 1997.

[3] The spread of counterfeiting: Knock-offs catch on. *The Economist*, 2010.

[4] S. Weis, S. Sarma, R. Rivest, and D. Engels. Security and privacy aspects of low-cost radio frequency identification systems. *Security in pervasive computing*, pages 50–59, 2004.

[5] L. Lu, J. Han, L. Hu, Y. Liu, and L.M. Ni. Dynamic key-updating: Privacy-preserving authentication for rfid systems. In *Proc. of IEEE PERCOM*, 2007.

[6] L. Lu, J. Han, R. Xiao, and Y. Liu. Action: breaking the privacy barrier for rfid systems. In *Proc. of IEEE INFOCOM*, 2009.

[7] L. Yang, J. Han, Y. Qi, and Y. Liu. Identification-free batch authentication for rfid tags. In *Proc. of IEEE ICNP*, 2010.

[8] L. Bolotnyy and G. Robins. Physically unclonable function-based security and privacy in rfid systems. In *Proc. of IEEE PERCOM*, 2007.

[9] Y.K. Lee, L. Batina, and I. Verbauwhede. Ec-rac (ecdlp based randomized access control): Provably secure rfid authentication protocol. In *Proc. of IEEE RFID*, 2008.

[10] Y.K. Lee, L. Batina, and I. Verbauwhede. Untraceable rfid authentication protocols: Revision of ec-rac. In *Proc. of IEEE RFID*, 2009.

[11] L.G. Roberts. Aloha packet system with and without slots and capture. *ACM SIGCOMM Computer Communication Review*, 5(2):28–42, 1975.

[12] B. Sheng, Q. Li, and W. Mao. Efficient continuous scanning in rfid systems. In *Proc. of IEEE INFOCOM*, 2010.

[13] D. Benedetti, G. Maselli, and C. Petrioli. Fast identification of mobile rfid tags. In *Proc. of IEEE MASS*, 2012.

[14] T.F. La Porta, G. Maselli, and C. Petrioli. Anticollision protocols for single-reader rfid systems: Temporal analysis and optimization. *IEEE Transactions on Mobile Computing*, 10(2):267 –279, 2011.

[15] R. Kumar, T.F. La Porta, G. Maselli, and C. Petrioli. Interference cancellation-based rfid tags identification. In *Proc. of ACM MSWiM*, 2011.

[16] J. Capetanakis. Tree algorithms for packet broadcast channels. *IEEE Transactions on Information Theory*, 25(5):505–515, 1979.

[17] J. Myung and W. Lee. Adaptive splitting protocols for rfid tag collision arbitration. In *Proc. of ACM MobiHoc*, 2006.

[18] Epcglobal radio-frequency identity protocols class-1 generation-2 uhf rfid protocol for communications at 860 mhz-960mhz, 2008.

[19] Information technology radio frequency identification for item management part 6: Parameters for air interface communications at 860 mhz to 960 mhz, 2010.

[20] T. Dimitriou. A secure and efficient rfid protocol that could make big brother (partially) obsolete. In *Proc. of IEEE PERCOM*, 2006.

[21] Li Lu, Yunhao Liu, and Xiang-Yang Li. Refresh: Weak privacy model for rfid systems. In *Proc. of IEEE INFOCOM*, 2010.

[22] M. Kodialam and T. Nandagopal. Fast and reliable estimation schemes in rfid systems. In *Proc. of ACM MobiCom*, 2006.

[23] C. Qian, H. Ngan, Y. Liu, and L.M. Ni. Cardinality estimation for large-scale rfid systems. *IEEE Transactions on Parallel and Distributed Systems*, 22(9):1441–1454, 2011.

[24] Y. Zheng and M Li. Pet: Probabilistic estimating tree for large-scale rfid estimation. *IEEE Transactions on Mobile Computing*, 11(11):1763–1774, 2012.

[25] M. Shahzad and A. Liu. Every bit counts - fast and scalable rfid estimation. In *Proc. of ACM MobiCom*, 2012.

[26] T. Li, S. S. Wu, S. Chen, and M. C. K. Yang. Generalized energy-efficient algorithms for the rfid estimation problem. *IEEE/ACM Transactions on Networking*, 20(6):1978 –1990, 2012.

[27] Y. Zheng and M. Li. Zoe: Fast cardinality estimation for large-scale rfid systems. In *Proc. of IEEE INFOCOM*, 2013.

[28] P.H. Cole and D.C. Ranasinghe. Networked rfid systems. *Networked RFID Systems and Lightweight Cryptography*, pages 45–58, 2008.

[29] S. Devadas, E. Suh, S. Paral, R. Sowell, T. Ziola, and V. Khandelwal. Design and implementation of puf-based unclonable rfid ics for anti-counterfeiting and security applications. In *Proc. of IEEE RFID*, 2008.

[30] R. Zhang, Y. Liu, Y. Zhang, and J. Sun. Fast identification of the missing tags in a large rfid system. In *Proc. of IEEE SECON*, 2011.

[31] P. Flajolet and G. Nigel Martin. Probabilistic counting algorithms for data base applications. *Journal of computer and system sciences*, 31(2):182–209, 1985.

[32] R. Motwani and P. Raghavan. *Randomized algorithms*. Chapman & Hall/CRC, 2010.

[33] P.S. Bullen. *Handbook of Means and their Inequalities*. Springer, 2003.

[34] Open rfid lab, http://pdcc.ntu.edu.sg/wands/orl.